DOCUMENT RESUME

ED 229 408 TM 830 305

AUTHOR Shiffler, Nancy L.
TITLE Evaluating Individualized Programs.
SPONS AGENCY Office of Special Education (ED), Washington, D.C.
PUB DATE Apr 83
CONTRACT G007903009
NOTE 27p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (120) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Evaluation Criteria; Individualized Instruction; *Individualized Programs; Intervention; *Outcomes of Education; *Program Effectiveness; *Program Evaluation; Special Programs
IDENTIFIERS *Causal Inferences; Intervention By Prescription Project

ABSTRACT

Several current approaches to evaluation stress the importance of making explicit the conceptual structure of a program--displaying the hierarchically related transactions which explain the program's movement from a set of "inputs" to a set of desired outcomes. However, individualized programs present particular difficulties in establishing this line of causality. When the program is the same for all individuals, everyone receiving the same intervention, then the conceptual structure is also the same for everyone--a common ground exists for the nature of the intervention, the outcomes, and the variables. With individualized programs, the specific interventions, outcomes, and rationale may vary from individual to individual. The common ground is now moved to a different level--the program has a common process for generating a conceptual structure for an intervention with each individual case and a common interest in how well the goals of each intervention are met. As a result, the evaluation of the program as a whole must try to integrate the findings from a set of disparate "programs," all having disparate measures of outcome goals. This paper attempts to define some of the difficulties inherent in these circumstances and to describe their impact on the evaluation for a particular individualized program. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED229408

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

EVALUATING INDIVIDUALIZED PROGRAMS

by

Nancy L. Shiffler

Presented at the Annual Meetings of the
American Educational Research Association,
Montreal, Quebec, Canada, April, 1983.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

N. L. Shiffler

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The research reported herein was performed
pursuant to Contract No. G007903009,
Office of Special Education, U. S. Department
of Education.

TM 830 305

ABSTRACT

Several current approaches to evaluation stress the importance of making explicit the conceptual structure of a program -- displaying the hierarchically related transactions which explain the program's movement from a set of "inputs" to a set of desired outcomes. However, individualized programs present particular difficulties in establishing this line of causality. When the program is the same for all individuals, everyone receiving the same intervention, then the conceptual structure is also the same for everyone -- a common ground exists for the nature of the intervention, the outcomes, and the variables. With individualized programs, the specific interventions, outcomes, and rationale may vary from individual to individual. The common ground is now moved to a different level -- the program has a common process for generating a conceptual structure for an intervention with each individual case and a common interest in how well the goals of each intervention are met. As a result, the evaluation of the program as a whole must try to integrate the findings from a set of disparate "programs," all having disparate measures of outcome goals. This paper attempts to define some of the difficulties inherent in these circumstances and to describe their impact on the evaluation for a particular individualized program.

EVALUATING INDIVIDUALIZED PROGRAMS

The aim of most program evaluation is to determine the extent to which a particular program or policy has had its intended effect upon a set of people or events. Inherent in this evaluation process is the need not only to measure change on a set of outcome variables but also to demonstrate that the program was in fact implemented and that its presence produced the change -- that is, "to establish a reasonable line of causality showing that a change occurred due to the presence of an implemented program" (Leinhardt, 1977). However, individualized programs, such as the individualized educational plans mandated for special education, present particular difficulties to the evaluator seeking to establish this line of causality. This paper attempts to define some of these difficulties and to describe their impact on the evaluation design for a particular individualized program.

In response to the need to identify the causal linkages in educational programs, considerable work has been done lately in developing evaluation models which provide for detailed specification of the process of the program (see for example, Borich, 1977; Cooley, 1978; Collet and Pintrich, 1979, 1980; Leinhardt, 1980). These models stress the importance of making explicit the conceptual structure of a programm --displaying the hierarchically, or sequentially, related transactions which explain how and why the program is expected to be able to move from a set of

"inputs" to a set of desired outcomes. The Conceptually Structured Program Evaluation model (Collet and Pintrich, 1979, 1980), in particualr, requires that these transactions be represented "in terms of changes in the state of learning (knowledge, skills, attitudes, etc.) of the target population." The intent of this approach is to focus the evaluation at the individual level rather than the program level, encouraging greater attention to individual differences.

When the program is the same for all individuals, that is, when everyone gets essentially the same intervention, then the conceptual structure is also the same for everyone. Individual differences in outcomes may be related to differences in entry level "skills," individual/treatment interactions, or differences in program implementation for each individual, but a common ground exists for the nature of the intervention, the intended outcomes, and the variables to be measured.

With individualized programs, however, the specific interventions, outcomes and reasons <u>for</u> the interventions and outcomes may vary from individual to individual. Thus, the conceptual structure varies for each case, as do the variables to be measured. The common ground is now moved up to a different level -- the program may (or should, if it is truly to be called a "program") have a common process for <u>generating</u> a conceptual structure for an intervention with each individual case and a common interest in assessing how

well the goals of each intervention are met. The evaluation of the program as a whole becomes rather like a metaevaluation in that one is trying to integrate the findings from a set of disparate "programs," all of which may have disparate measures of outcome goals; in this case, all the "programs" are essentially N of 1 studies.

This way of looking at the evaluation of individualized programs has two primary implications for evaluation design. First, the assessment of degree of program implementation must include an assessment of the extent to which the program's model for generating a conceptual structure and intervention was used, as well as an assessment of whether the specific interventions were implemented. As a result, the "reasonable line of causality" becomes much longer, hence harder to establish; there are now more points at which implementation failure can occur and thus, more points at which implementation must be measured.

Second, program success can occur in a variety of ways for different cases, presenting considerable methodological problems. Bryk and Light (1981) define the methodological difficulty as follows:

> A highly individualized program can be effective without all of its subjects moving in a particular direction on all dimensions within a single evaluation time frame. Yet this uniformity constitutes the implicit assumption of all traditional univariate and multivariate analysis methods. In individualized programs the search for mean differences across groups, variable by variable, is often futile. There is usually considerably less statistical power than might appear on the basis of total sample size since each outcome variable is only relevant for a small subset of cases at any particular point in time (p. 20).

Bryk and Light go on to advocate the use of approaches from N of 1 research, such as multiple baseline designs or growth curve models, but they suggest that these approaches also have problems when applied to highly individualized programs. In particular, they point out that the degree of control over the intervention (over its onset, for example) required by these techniques is usually not present; rather, the control is "shared" interactively with the child and with the environment. In addition, Bryk and Light point out that these designs assume that there are only a limited number of variables of interest rather than the wide variety that may appear in individualized plans, and in any case the currently available metrics are adequate for only a rather narrow range of behaviors.

Thus, evaluators of individualized programs are confronted with the task of evaluating a complex array of interventions and intended outcomes with a methodology which is not yet totally adequate for the task. If one is interested in making general statements about program effectiveness, then some means must be found for aggregating results across cases. Measuring "all possible outcome variables" is unwieldy, while choosing a small set of "marker" variables, as suggested by Bryk and Light, may risk shortchanging some unique interventions. Global measures of success such as Goal Attainment Scaling (Kiresuk and Sherman, 1978) avoid this problem by using degree of goal attainment, regardless of the goal, as the common metric.

This approach, however, requires a reasonable amount of agreement among goal-setters as to what constitutes an "expected" level of attainment for an intervention and agreement about the relative "importance" or "distance" ratings which should be assigned to each goal. Without this agreement, Goal Attainment Scaling does not establish the common metric it is supposed to provide.

While none of these approaches alone is satisfactory, some benefit may be gained by using two or more and playing the results off each other. The remainder of this paper describes an evaluation design which attempted to do just that. Although the results do not provide a definite solution to the problems of evaluating individualized programs, they do suggest some directions which might help confront these problems.

The Intervention By Prescription Project

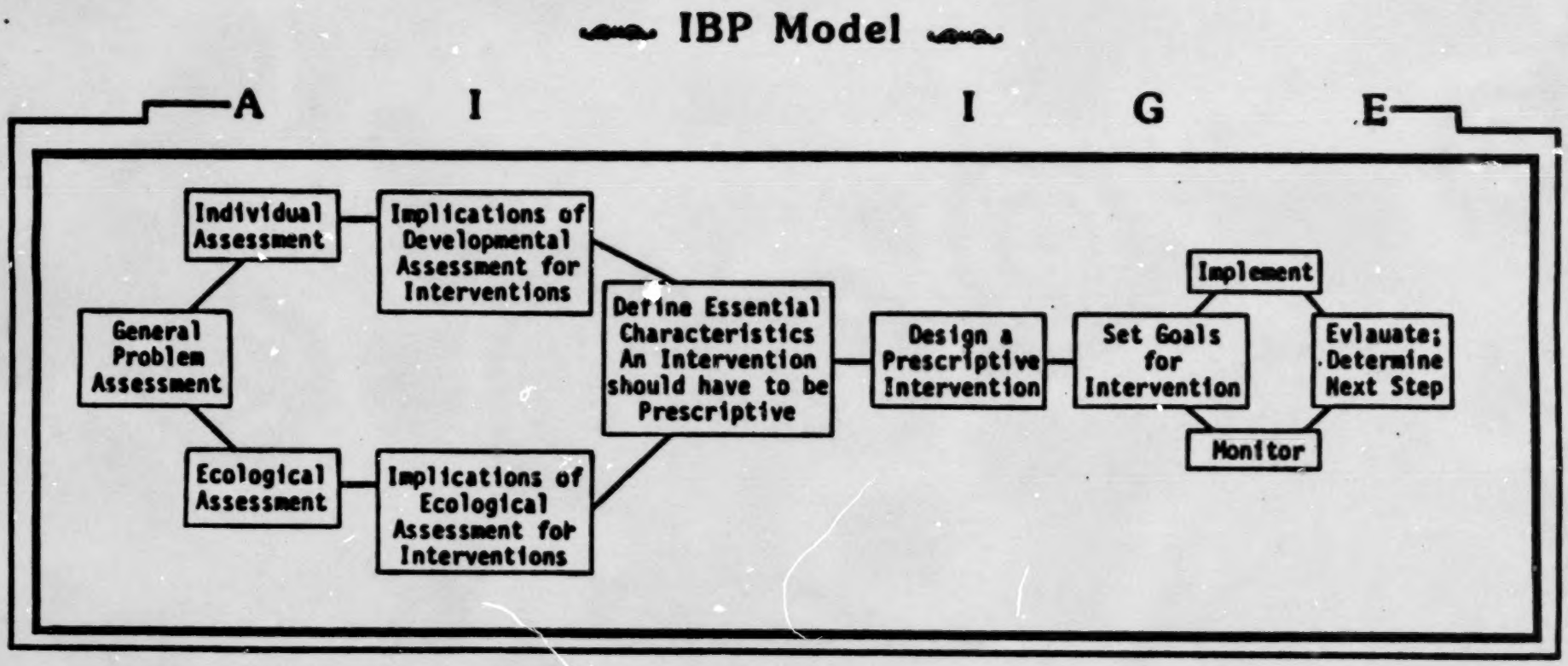
The Intervention By Prescription (IBP) project was a federally funded model demonstration project which sought to develop and test a diagnostic/intervention model for school-aged children with emotional problems. The project grew out of a concern that special education guidelines, as they are currently interpreted, have led to an emphasis on labelling and placement rather than on problem analysis and prescriptive interventions (Rezmierski and Rubinstein, 1982). Diagnostic personnel have been increasingly tied to ritualistic procedures with heavy reliance on psychometrics.

The accumulation of information from a number of unrelated sources is frequently inadequate for the task of translating diagnostic information into an appropriate intervention program.

A premise of the IBP project model is that developmental and ecological theory can provide constructs around which more systematic information gathering can occur, producing information which will lead more directly to prescriptive interventions. The model has identified common constructs from the theories which seem to be most pertinent to the diagnosis of emotional problems within school-aged children; these constructs include impluse control, affective preoccupation, social understanding (perspective-taking) and conflict resolution, in the developmental area, and the consideration of the negative and positive supports from the home, school, and peer group ecologies.

From this assessment, practitioners draw out the implications, outlining the essential characteristics of a prescriptive intervention. The details of the intervention are worked out, including a set of realistic goals for, or predictions about, what changes should occur if the intervention is on target. Finally a monitoring plan is worked out for determining the success of the intervention. A diagram of the overall operation of the model appears in Figure 1.

FIGURE 1:


IBP Model
A
I
I
G
E
Individual Assessment
General Problem Assessment
Ecological Assessment
Implications of Developmental Assessment for Interventions
Implications of Ecological Assessment for Interventions
Define Essential Characteristics An Intervention should have to be Prescriptive
Design a Prescriptive Intervention
Set Goals for Intervention
Implement
Monitor
Evlauate; Determine Next Step

The steps of the model can be thought of as a process for generating the conceptual framework for an intervention plan which is responsive to the developmental and ecological issues of the problem. The framework provides the rationale for why and how the intervention should work. The selection of goals and measures for monitoring the progress of the intervention should follow from this rationale.

The IBP model was developed, implemented, and evaluated over a three-year period in eight elementary and junior high schools in three school districts in Michigan. During this time the model was used to analyze and intervene with 184 child/school problems referred to child study teams by classroom teachers.

The Evaluation of the Project

The IBP project undertook to develop a diagnostic/intervention model, train a set of practitioners in its use, and implement the model with a group of children. Accordingly, an evaluation component was developed for each of these steps -- validating the model against the developmental theories from which it was developed, evaluating the degree to which training in the reliable use of the model could be carried out, and evaluating the model's effectiveness in actual practice. However, the fact that the steps described above were carried out, if not simultaneously, then at least with considerable overlap in time posed some difficulties for evaluation. The evaluation

design for the project was shaped by the following considerations:

1) The IBP model was in the process of being developed over the course of the project; a good evaluation design, therefore, would have to be flexible enough to accommodate and contribute to changes in the model. As with any model in its developmental phase, the key variables and dynamics of the process were still being defined and refined, thus, the evaluation would need to emphasize an exploratory, descriptive approach much more than a tightly defined experimental design.
2) The model was a process for arriving at an intervention for a problem -- it was not simply the intervention itself; thus, it would be particularly important for the evaluation to capture the crucial points in the process leading to the intervention.
3) The interventions produced by the model were individualized -- the problems, interventions, and expected outcomes differed for each case. Consequently, no single outcome measure would be applicable in every case, and restricting the possible measures would, in fact, contradict the very intent of the model. The evaluation design, therefore, would have to allow for a variety of measures of problem change, while still providing a means for aggregating case results.

Given these considerations, the primary dilemma for the evaluation design was to maintain the flexibility and open-endedness needed for a developing, process-oriented model while at the same time structuring the evaluation sufficiently to allow for the combining of information across cases and across child study teams for the eight schools. The source of structure for this project was the IBP model itself; the steps of the process -- Assessment, Implications, Prescription/Intervention, Goal-setting, and Evaluation -- were common (or should have been when the model was being implemented properly) to each case, even though the problems, the interventions, and the means for monitoring them might differ.

The IBP model defines a process for conceptualizing, designing, and evaluating an individually prescriptive intervention. Accordingly, the evaluation of the model began by asking whether the model "worked" for an individual case: 1) was the project model used to develop the intervention for this case; 2) was the intervention implemented; and 3) did the intervention work as expected. The primary vehicle for gathering information to answer these questions was a partially open-ended set of worksheets, referred to as a Problem Analysis record, completed for each case by the project's primary contact person in each school building. The worksheets required summary entries for each phase of the model diagrammed in Figure 1:

1) assessment results were summarized, stating the child's level of operation in the four developmental areas and the strengths and weaknesses in the child's environment;
2) the implications of the assessment were spelled out, in particualar by defining the essential characteristics an intervention would need to have to be prescriptive for the child's needs;
3) the intervention itself was described -- its structure, rationale, and duration;
4) using a simplified version of Goal Attainment Scaling, predictions about possible levels of success (from "most unfavorable," to "expected level," to "best anticipated") were stated;
5) the monitoring plan and results were summarized.

Thus, event though the content of each case and each intervention varied, parallel process information was available for each case.

These worksheets served as the primary, but not the sole, data source for the evaluation of both model implementation and model effectiveness. Additional measures of child change were provided by a concurrent study of the prevalence of emotional stress in school-aged children, which the IBP project was carrying out in these and other local schools. Teachers were asked to 1) identify children in their classes whom they felt were experiencing "emotional stress;" 2) rate the urgency of the problems; and 3) fill

out a Hahnemann school behavior rating form (Swift and Spivack, 1975) on each identified child. These teacher ratings were collected four times over the three-year period, providing a set of "pre-post" measures for the children actually antervened with (these children were drawn from a stratified random sample of this pool of "stressed" children).

Evaluating Implementation

The major implementation concern for the project was whether or not the project model was being used to develop the rationale and structure for the interventions designed for the target children. In particular, it was important to determine whether the interventions were consistent with the developmental levels of the children, as ascertained during the assessment phase of the process. Accordingly, the procedures for analyzing the case records focused on the five steps of the model (assessment, implications, interventions, goal-setting, evaluation) -- their presence, their quality, and their linkage to adjoining steps. The project staff developed a coding procedure for a structured content analysis of the case records for each of these components.

The coding of the presence of a step was straightforward, as was the coding of quality, once the basic criteria were set. Linkages between steps, however, were much more difficult to measure and code. This occurred, in part, because the process for making these

linkages was defined only as the model evolved over the course of the project and, thus, was not really made explicit to the field staff. In addition, the linkages are themselves processes rather than endpoints, requiring an inference that, for example, the implications recorded on a form were, in fact, drawn from the asessment data. If the correspondence between assessment and implication is one-to-one, i.e., if only one possible implication for intervening can be drawn from a particular assessment level, then the validity of the inference may be relatively easy to judge. If, however, the correspondence is one-to-many or many-to-one, then the task becomes more difficult and introduces the question of whether it is more important that the implications be as complete as possible or that they include the most "crucial" implications for that particular child.

These definitional and measurement issues only began to be developed within the time period of the project and are areas in need of further research. For purposes of the current evaluation, we settled on a rough measure of linkage. We reasoned that if the implications and the interventions could be judged to be consistent with particular assessment levels (or a range of levels), and these judgments in fact matched the actual assessment, then it might be plausible to infer that the linkage between the steps of the model was occurring. This procedure should be viewed only as an initial attempt to cope with the complex measurement problem presented by the linkage process.

These "backward linkage" judgments for the implications and intervention steps were made using a blind coding procedure -- raters were looking at only the summarized implications or interventions, from which they attempted to reconstruct the assessment levels. These reconstructions were then compared to the actual assessment levels.

In addition to these blind coding steps, two other important linkages were assessed. First, the developmental assessments were judged for consistency between the developmental levels given and the supporting data used to back up the assessment. Secondly, the goals (and their accompanying measures) for each intervention were rated on the extent to which they were directly related to the intervention.'

Thus, this portion of the evaluation was an attempt to capture the crucial first steps of the long "line of causality" for assessing the effects of the project model -- was the project model being used to design the interventions. As noted earlier, given the evolving nature of the project model (and, hence, the evaluation design), these measures could only serve as rough markers of implementation.

'For example, for an intervention focused on classroom disruptions, reduction of out of seat behavior, talking out, and so forth is a directly related goal; improved academic performance, while potentially an outcome of such an intervention, would not be directly related because the interim goals of reduced classroom disruption and, perhaps, increased task completion would probably have to be observed first. A goal attainment scale which included all three of these steps, however, would be rated as directly related.

The analysis of the implementation data for all cases indicated that implementation improved over time and in a pattern that was generally consistent with the timing of the project's training and development efforts. For example, much of the training in the assessment process occurred prior to or at the very beginning of the time period for which case data were being collected; accordingly, the ratings of agreement between the assessments and their accompanying backup data tended to remain relatively stable and relatively high (about 80% agreement overall) over time. The implications process, on the other hand, was probably the most difficult step to define and exemplify, with much of the training occurring at the beginning of the final project year. Consequently, the measures of implementation of the implications phase, including linkage to the assessment, showed considerable improvement in the final year, although they did not achieve a particularly high level of implementation.

While implementation tended to improve over time for each of the model components, only a few cases had all components completely implemented. This is not surprising given that the project model was still being developed; school staff had no prototypes of fully implemented cases to work from and, indeed, were involved in developing prototypes. However, given this level of implementation, one would not expect to make any definitive statements about the model's ability to effect child change; the line of

causality is only tenuous at this time. On the other hand, it is possible within these limitations to explore the relationship between degree of implementation and intervention success and to examine the issues involved in measuring child change.

Evaluating Child Change

As indicated earlier in this paper, the project's general approach to evaluating child change was that no one method could provide adequate information when interventions were as highly individualized as those in the project. With predefined outcome measures, not all cases could be expected to change on each variable because individual interventions emphasize different areas of needed change. With more global goal attainment measures, the "distances" between levels of attainment might vary considerably across interventions depending on how ambitious the intervention was. By combining methods, we hoped to use the information from one approach as a means for interpreting the information obtained from the other.

Beginning with a global indicator of success -- to what extent did the interventions meet their goals -- we looked at this measure from two directions: 1) its relationship to level of implementation of the project model; and 2) the effects of goal attainment on a broader set of outcome variables.

Of all the interventions completed during the course of the project, 82% reached their goals at the expected level

or better. In order to better describe the actual distance moved, we had rated each intervention, prior to analysis, on whether it was primarily "symptom-focused," i.e., narrowly focused on a small piece of behavior, or "problem-focused," i.e., focused on the broader scope of the problem. Over all, about 65% of the interventions were problem-focused, with interventions in the final project year being significantly more problem-focused than those in the first year of the model's implementation.

While problem-focused and symptom-focused interventions had equal success rates, their differential effects appeared in the relationship between model implementation and success rate. Using the extent to which an intervention was consistent with the child's developmental assessment as a measure of implementation, we found that for problem-focused interventions, those that were on target were slightly (but not significantly) more likely to be successful; for symptom-focused interventions, those that were off target were actually somewhat more likely to be successful. While these trends are hardly conclusive, they suggest the possibility that the project model (i.e., designing developmentally-related interventions) may have the greatest effect when one is responding to the broader scope of a problem rather than focusing on fairly limited gains.

Regardless of the role of the project model, to what extent did successfully attaining the goals of an intervention carry over to other measures of change? Our

first source of information on this point was another global measure of improvement -- the classroom teachers' pre and post ratings of the urgency of the problem. While they started out at about the same level of urgency, the problems with unsuccessful interventions were perceived as more urgent following the intervention, and those with successful interventions were perceived as less urgent (gain score $t_{(131)}=3.3$, $p<.05$). This result indicates, at the least, some consistency in the perceptions of the project field staff and the classroom teachers concerning improvement of the problem.

As a final means of exploring the effects of the inteventions, we used the classroom teachers' pre-post ratings of the children on the Hahnemann Elementary[2] School Behavior Rating Scale. This 60-item scale is grouped into 15 behavioral factors[3] (plus a rating of academic achievement) which the authors of the scale found to be related to successful classroom performance

The analysis of the pre-post results across all cases with "successful" interventions showed no significant gain over time. These results are consistent with the point made by Bryk and Light (1981), discussed earlier, that with

---

[2]Because of the small number of cases, junior high results are not reported here.

[3]Originality, Independent Learning, Involvement in Class, Productivity with Peers, Intellectual Dependence, Failure Anxiety, Unreflective/Impulsive, Irrelevant Talk, Social Over-involvement, Negative Feelings, Holding Back/ Withdrawn, Critical/Competitive, Blaming, Approach to the Teacher, Inattentiveness.

individualized programs one does not expect unidirectional movement on all variables for all cases. Consequently, we reasoned that only by looking more specifically at the structure and targets of the interventions could we gauge their effects on these variables.

Our early content analysis of the interventions indicated that they could be clustered roughly into 12 "target behavior" groups, such as task completion, expression of feelings, social skills improvement, responsiveness/participation, anxiety, disruptive classroom behavior, and so forth. Thus, prior to any further analysis of the Hahnemann factors, we attempted to identify for each intervention type a constellation of factors which those interventions might be expected to affect. The profiles of the factor gain scores for each intervention type, therefore, ought to show peaks for those factors judged to be most sensitive to that particular type of intervention. The small group sizes which resulted from breaking the interventions down into these 12 clusters precluded any valid significance testing; consequently, the results reported here are strictly descriptive.

The profiles for five of the twelve intervention types showed a reasonable match with our predicted constellation of factors: Responsiveness/Participation, Motivation/ Attitude, Social Relations (i.e., developing friendships), Task Completion, and Anxiety. For example, children with interventions targetted towards responsiveness and greater

participation in class showed, as a group, more involvement in class, less withdrawn behavior, less inattentiveness, and more contact with the teacher.

These results suggest that successful interventions of these types were most likely to carry over into the classroom. Given that the Hahnemann scales are based on teacher perceptions, the results may also indicate that these target behaviors were ones for which teachers were more willing to recognize change, compared to, for example, disruptive classroom behavior, which may evoke a more rigid response from a teacher.

Other bases for clustering interventions might also be useful. For example, interventions carried out outside of the classroom -- such as individual counselling -- might be expected to show less carry over into classroom behavior (or teacher perception of classroom behavior) than would interventions carried out in the classroom. Coupled with measures of time, this sort of analysis could provide information about the time lag to be expected for some types of interventions to show results. Obviously, the more these issues can be identified a priori, the stronger the conclusions that can be drawn from them. This could become increasingly possible as the development of the project model continues.

Conclusion

This study demonstrated two of the major difficulties in evaluating individualized programs. First, the "line of causality" for demonstrating the program's effect on outcome is extended considerably, requiring more measurement points and frequently resulting in less control over the entire implementation process. It becomes increasingly important under these circumstances to identify the most crucial links in the chain of causality and to focus program control and measurement efforts on these points.

Secondly, the variety of possible outcomes for individualized interventions can easily fragment efforts to assess program effectiveness. The results reported here suggest the value of clustering interventions as a middle ground between the total aggregation of cases and a series of N of 1 analyses. It may capture results which might otherwise be missed, and it adds descriptive power to more global measures. However, one is still left with the loss of statistical power which results from subdividing the total sample. Only by linking this approach to more global analyses of goal attainment (provided the goal-defining process is carefully monitored) can some of this power be retained.

The evaluation of individualized programs is never likely to be straightforward -- there are simply too many variables to be controlled. The approaches described here, however, suggest that there are ways of structuring the

evaluation to take better advantage of the controls that are there.

REFERENCES

Borich, Gary, Program evaluation: New concepts, new methods. Focus on Exceptional Children, May, 1977, 9(3), 1-14.

Bryk, Anthony S. and Light, Richard J., Designing evaluations for different program environments, in Educational Evaluation Methodology: The State of the Art (Ronald A. Berk, ed.). Baltimore:Johns Hopkins Press, 1981.

Collet, LeVerne S. and Pintrich, Paul, Towards a conceptually structured program evaluation model. Paper presented at the Fifth Annual Evaluation Network Conference. Cincinnati, Ohio, 1979.

Collet, LeVerne S. and Pintrich, Paul, Conceptually structured program evaluation: A heuristic model for designing program evaluations. Paper presented at the Annual Meetings of the American Educational Research Association, Boston, April, 1980.

Cooley, William, Explanatory observational studies. Educational Researcher, 1978, 7(9), 9-15.

Kiresuk, Thomas J. and Sherman, Robert E., Goal attainment scaling: A general method for evaluating comprehensive community mental health programs. Community Mental Health Journal, 1978, 4, 443-453.

Leinhardt, Gaea, Program evaluation: An empirical study of individualized instruction. American Educational Research Journal, 1977, 14(3), 277-293.

Leinhardt, Gaea, Modeling and measuring educational treatment in evaluation. Review of Educational Research, 1980, 50(3),393-420.

Rezmierski, Virginia and Rubinstein, Marla Frudden. To punish or to heal: The issues and dynamics of educating emotionally disturbed children, in Iowa Monograph: Current Issues in Behavior Disorders, (Carl R. Smith and Barbara J. Wilcox, eds.). DeMoines, Iowa: Iowa Department of Public Instruction, August, 1982.

Spivack, George and Swift, Marshall, Hahnemann Elementary School Behavior Rating Scale (HESB) Manual. Hahnemann Community Mental Health/Mental Retardation Center, Department of Mental Health Sciences, Hahnemann Medical College and Hospital, Philadelphia, Pa., 1975.